Breast Cancer Research and Treatment (2005) 90: 127–137

Report

Distinct breast cancer incidence and prognostic patterns in the NCI's SEER program: suggesting a possible link between etiology and outcome

William F. Anderson[1], Ismail Jatoi[2], and Susan S. Devesa[3]
[1]*DHHS/NIH/NCI/Division of Cancer Prevention, EPN, Bethesda, MD, USA;* [2]*Department of Surgery National Naval Medical Center, Uniformed Services University, Bethesda, MD, USA;* [3]*DHHS/NIH/NCI/Division of Cancer Epidemiology and Genetics, Bethesda, MD, USA*

Key words: actuarial survival, age frequency distribution, breast cancer age-specific incidence rates, etiology, hazard rates, prognosis

Summary

Background. Breast cancer is a heterogeneous and chronic disease with relapses and death occurring 25 years or more after primary diagnosis. Standard tumor characteristics are used to predict initial relapse or death, but their ability to estimate long-term patterns of failure may be limited.

Methods. To further evaluate the significance of standard tumor features, we compared incidence and prognostic patterns in the National Cancer Institute's (NCI's) large-scale population-based Surveillance, Epidemiology, and End Results (SEER) program for high-risk versus low-risk breast cancers, i.e., size >2.0 versus ≤ 2.0 cm, lymph node positive versus negative, high versus low histologic grade, and hormone receptor negative versus positive expression, respectively. Data were stratified by age 50 years to approximate menopause.

Results. High-risk versus low-risk breast cancers demonstrated two very different incidence and prognostic patterns. Age-specific incidence rates among women with high-risk tumors increased until age 50 years then flattened, whereas rates among women with low-risk tumors increased continuously with aging. Hazard rates for breast cancer death spiked sharply two years following primary breast cancer diagnosis among women with high-risk but not with low-risk tumors. Paradoxically, hazard function crossed over 6–8 years following breast cancer diagnosis, with hazard rates lower for high-risk than for low-risk breast cancers.

Conclusion. Distinct incidence and prognostic patterns among high-risk and low-risk breast cancers suggest a possible link between breast cancer etiology and outcome. These epidemiologic results appear to complement emerging molecular genetic techniques, showing distinct genotypes for high-risk and low-risk breast cancer phenotypes.

Introduction

Given the general view of breast cancer as a biologically heterogeneous and chronic disease process, clinicians have relied upon incident tumor characteristics to estimate high-risk and low-risk patterns of relapse and/or death. High-risk characteristics such as early age-at-onset, tumor size >2.0 cm, axillary lymph node (LN) positive, high histologic grade, estrogen receptor (ER) negative expression, and progesterone receptor (PR) negative expression are associated with relatively unfavorable prognosis and early relapse. On the other hand, low-risk characteristics such as later age-at-onset, size ≤ 2.0 cm in diameter, LN negative status, low histologic grade, ER positive expression, and PR positive expression are correlated with relatively favorable prognosis.

While these tumor features may predict near-term prognosis, their ability to forecast long-term patterns of failure is suspect. For example, Hilsenbeck et al. demonstrated crossover effects for ER expression after 3–5 years of follow-up [1]. Before 3 years, the risk of relapse was greater for ER negative than for ER positive breast cancers. At 3 years, risk of relapse was neutral for ER expression. After 5 years, risk of relapse was greater for ER positive than for ER negative breast cancers.

Some clinical evidence even suggests that high-risk and low-risk tumor characteristics may be predestined prior to primary breast cancer diagnosis [2–4]. Recent molecular studies seemingly support this view, demonstrating distinct gene expression signatures for high-risk and low-risk breast cancer populations [5, 6]. However, these molecular results have not been confirmed in a population-based setting.

To further explore the relationship between initial tumor features and prognosis, we examined age-specific incidence rates, age frequency distribution at diagnosis, actuarial breast cancer survival, and hazard function among women with high-risk versus low-risk breast cancers in the large-scale population-based Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute (NCI).

Materials and methods

We used the NCI's SEER Cancer Incidence Public-Use database (November 2003 submission) to analyze invasive female breast cancers [7]. The SEER database included overlapping 9 and 12 Registry Databases. The 9 Registry Database collected data for the years 1973–2001 from SEER's original catchment regions, including registries in Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah. The 12 Registry Database collected information for the years 1992–2001 from SEER's original 9 registries plus Los Angeles, San Jose-Monterey, and the Alaskan Native Tumor registries. Although operative for fewer years than the 9 Registry Database, the 12 Registry Database provided more detailed information for tumor characteristics and hormone receptor expression. For example, SEER did not collect information concerning tumor size, axillary lymph nodal (LN) status, and histologic grade until 1988, and did not collect data regarding hormone receptor expression until 1990. Given that this analysis required data for tumor characteristics, breast cancer cases were extracted from SEER's 12 Registry Database (1992–2001). We adopted age 50 years as our surrogate measure for menopause [8].

Tumor characteristics were categorized as high-risk or low-risk. High-risk or low-risk characteristics were tumor size >2.0 or ≤ 2.0 cm in diameter, positive or negative axillary lymph nodes (LN positive or negative), high or low histologic grade, ER negative or positive expression, and PR negative or positive expression, respectively. We also created composite high-risk or low-risk groups for those breast cancers with exclusively unfavorable or favorable tumor characteristics, respectively. For example, composite high-risk breast cancers were >2.0 cm in diameter + LN positive + high grade + ER negative expression. Composite low-risk breast cancers were ≤ 2.0 cm + LN negative + low grade + ER positive expression. To conserve sample size, PR expression was excluded from these composite definitions.

Incidence patterns

Incidence rates with standard errors (SE) were calculated using SEER stat 5.2.2, age-adjusted to the 2000 US standard, and expressed per 100,000 woman-years. Relative risks for high-risk compared to low-risk tumor characteristics were expressed as incidence rate ratios (RR), where a high-risk characteristic was compared to a low-risk characteristic with an assigned RR of 1.0. RRs were tested by calculating approximate 95% confidence intervals [9]. Age-specific incidence rate curves were charted on a log–log scale, as originally described by Armitage and Doll [10, 11]. Probability density function curves for age at diagnosis were plotted, as previously described [12, 13]. In brief, the probability density function reflected a smoothed age frequency distribution at the time of primary breast cancer diagnosis.

Prognostic patterns

We used the Kaplan–Meier (KM) product-limit method to estimate actuarial breast cancer survival [14]. Cumulative survival curves for high-risk and low-risk characteristics were compared with the two-sided log rank test [15]. Cox proportional-hazards model was used to examine the relative effects of tumor characteristics upon breast cancer survival [16]. SEER stat 5.2.2 calculated the hazard function for breast cancer death as relative hazard rates, derived from life-table estimates for expected survival. The hazard death rate was a conditional rate of breast cancer death in a specified time interval following breast cancer diagnosis given that the subject was alive at the beginning of that time interval. Hazard rate described not only the rate of failure over time, but also the relative magnitude of failure. The relative rate of failure was plotted on the y-axis, and the time interval following breast cancer diagnosis was plotted on the x-axis.

Results

The SEER 12 Registry Database collected data for $n = 242{,}549$ invasive female breast cancer cases diagnosed during the years 1992–2001 (Table 1). There were 56,899 (23.5%) breast cancer cases age <50 years and 185,650 (76.5%) cases age ≥ 50 years. Median age-at-diagnosis was 62 years for all breast cancer cases combined. Median tumor size was 1.6 cm for all breast cancer cases, with larger tumors being more common among women <50 years (2.0 cm) than among women ≥ 50 years (1.5 cm). The overall incidence rate was 132.5 per 100,000 woman-years. Breast cancer incidence rates peaked among women ages 70–79 with the RR of 11.1 (95% CI, 10.9–11.2) compared to women age <50 years. The RR for black to white race was 0.9 (95% CI, 0.9–0.9) for all breast cancer cases, 1.0 (95% CI, 1.0–1.0) for younger women, and 0.8 (95% CI, 0.8–0.8) for older women.

All single and composite high-risk tumor characteristics compared to low-risk tumor characteristics were more common among women age <50 years than among women age ≥ 50 years (Table 1). Ninety-five percent confidence limits (95% CI) were very tight for all comparisons. For example, RR for large (>2.0 cm) compared to small (≤ 2.0 cm) tumor diameter was 0.7 (95% CI, 0.7–0.8) for younger women and 0.5 (95% CI, 0.5–0.5) for older women. Similar patterns were observed for LN status, histologic grade, ER, and PR. The group with 'unknown and other' data ranged from 10 to 25% for single tumor characteristics.

Table 1. Descriptive statistics for standard tumor characteristics in SEER's 12 registry Database among female breast cancer cases collected during 1992–2001

	All breast cancer cases combined						Women age < 50 years						Women age ≥ 50 years					
Sample size	242,549						56,899						185,650					
% of total cases	100.0%						23.5%						76.5%					
median age	62 years						44 years						68 years					
median tumor size	1.6 centimeters						2.0 centimeters						1.5 centimeters					
rate (*SE*)	132.5 (*0.3*)						42.6 (*0.2*)						368.0 (*0.9*)					
Variable	*N*	*%*	*Rate*	*RR*	*95% CI*		*N*	*%*	*Rate*	*RR*	*95% CI*		*N*	*%*	*Rate*	*RR*	*95% CI*	
Age (years)																		
< 50	56,899	23%	42.6	1.0	Lower	Upper												
50–59	51,756	21%	279.9	6.6	6.5	6.7												
60–69	51,982	21%	388.9	9.1	9.0	9.2												
70–79	51,957	21%	471.7	11.1	10.9	11.2												
80+	29,955	12%	433.0	10.2	10.0	10.3												
Race																		
White	202,824	84%	138.3	1.0	Lower	Upper	43,945	77%	43.1	1.0	Lower	Upper	158,879	86%	387.7	1.0	Lower	Upper
Black	20,320	8%	120.3	0.9	0.9	0.9	6,575	12%	43.8	1.0	1.0	1.0	13,745	7%	320.5	0.8	0.8	0.8
Other	18,053	7%					5,971	10%					12,082	7%				
Unknown	1,352	1%					408	1%					944	1%				
Tumor size																		
≤2.0 cm	142,021	59%	77.9	1.0	Lower	Upper	29,314	52%	22.1	1.0	Lower	Upper	112,707	61%	224.0	1.0	Lower	Upper
> 2.0 cm	76,511	32%	41.7	0.5	0.5	0.5	21,995	39%	16.4	0.7	0.7	0.8	54,516	29%	107.9	0.5	0.5	0.5
Unknown or other	24,017	10%					5,590						18,427					
Lymph nodes (LN)																		
LN negative	146,223	60%	80.1	1.0	Lower	Upper	30,713	54%	23.1	1.0	Lower	Upper	115,510	62%	229.3	1.0	Lower	Upper
LN positive	68,842	28%	37.9	0.5	0.5	0.5	21,673	38%	16.2	0.7	0.7	0.7	47,169	25%	94.8	0.4	0.4	0.4
Unknown or other	27,484	11%					4,513	8%					22,971	12%				
Histologic grade																		
Low	118,321	49%	64.8	1.0	Lower	Upper	23,013	40%	17.4	1.0	Lower	Upper	95,308	51%	188.9	1.0	Lower	Upper
High	78,805	32%	43.2	0.7	0.7	0.7	25,232	44%	18.7	1.1	1.1	1.1	53,573	29%	107.2	0.6	0.6	0.6
Other or Unknown	45,423	19%					8,654	15%					36,769	20%				

Table 1. (Continued)

	All breast cancer cases combined							Women age < 50 years							Women age ≥ 50 years						
Sample size	242,549							56,899							185,650						
% of total cases	100.0%							23.5%							76.5%						
median age	62 years							44 years							68 years						
median tumor size	1.6 centimeters							2.0 centimeters							1.5 centimeters						
rate (*SE*)	132.5 (*0.3*)							42.6 (*0.2*)							368.0 (*0.9*)						
Variable	*N*	*%*	*Rate*	*RR*	*95% CI*			*N*	*%*	*Rate*	*RR*	*95% CI*			*N*	*%*	*Rate*	*RR*	*95% CI*		
ER																					
ER positive	145,403	60%	79.5	1.0	Lower	Upper		29,821	52%	22.5	1.0	Lower	Upper		115,582	62%	229.0	1.00	Lower	Upper	
ER negative	43,586	18%	24.0	0.3	0.3	0.3		15,204	27%	11.3	0.5	0.5	0.5		28,382	15%	57.3	0.3	0.2	0.3	
Other or Unknown	53,560	22%						11,874	21%						41,686	22%					
PR																					
PR positive	121,698	50%	66.6	1.0	Lower	Upper		27,364	48%	20.6	1.0	Lower	Upper		94,334	51%	187.2	1.0	Lower	Upper	
PR negative	60,926	25%	33.4	0.5	0.5	0.5		16,420	29%	12.2	0.6	0.6	0.6		44,506	24%	88.9	0.5	0.5	0.5	
Other or Unknown	59,925	25%						13,115	23%						46,810	25%					
Composite profile																					
Low-risk	49,118	20%	27.0	1.0	Lower	Upper		7,946	14%	6.0	1.0	Lower	Upper		41,172	22%	81.8	1.0	Lower	Upper	
High-risk	6,615	3%	3.6	0.1	0.1	0.1		2,853	5%	2.1	0.3	0.3	0.4		3,762	2%	7.6	0.1	0.1	0.1	
Other or Unknown	186,816	77%						46,100	81%						140,716	76%					

Key: %, percent; Rate, age-adjusted (2000 US standard) incidence rate per 100,000 woman–years; RR, rate ratio where a high-risk characteristic is compared to a low-risk characteristic with an assigned RR of 1.0; 95% CI, 95% confidence interval with lower and upper bounds; Median age at diagnosis; Median tumor size at diagnosis; ER, estrogen receptor; PR, progesterone receptor; Composite low-risk prognostic factor profile, i.e., breast cancer cases with tumors ≤2.0 cm + LN negative + low grade + ER positive; Composite high-risk prognostic factor profile, i.e., breast cancer cases with tumors > 2.0 cm + LN positive + high grade + ER negative

RR for composite high-risk prognostic profiles compared to composite low-risk profiles was 0.3 (95% CI, 0.3–0.4) for younger women and 0.1 (95% CI, 0.1–0.1) for older women. There was a large amount of 'unknown and other data' for the composite group (77%, $n = 186,861$ of 242,549), which was actually composed mostly of *other* rather than *unknown* tumor characteristics. For example, where composite high-risk tumors would include breast cancers that were > 2.0 cm + LN positive + high grade + ER negative, *other* breast cancers for the composite designation might include breast cancers that were > 2.0 cm + LN positive + high grade + *ER positive*, or tumors that were > 2.0 cm + LN positive + *low grade* + ER negative, etc.

Figure 1 illustrates age-specific incidence rate curves for all breast cancer cases and single tumor characteristics. Rates for all cases increased rapidly until age 50 years, and then continued to rise at slower rates (Figures 1(a–f)). Rates were higher for Blacks than for Whites prior to age 50 years (Figure 1(a)), after which there was ethnic crossover with rates higher for Whites than for Blacks [17]. Rates for all single high-risk characteristics increased rapidly until age 50 years then rose at a much slower rate or flattened (Figures 1(b–f)). Rates for all single low-risk characteristics increased rapidly until age 50 years then continued to rise at a slower pace, similar to rates for all cases. We observed varying amounts of incidence rate crossovers for all high-risk and low-risk tumor characteristics (Figures 1(b–f)). For example, rates were higher for high grade tumors than for low grade tumors prior to age 50 years, after which rates were higher for low grade than for high grade tumors. Age-specific rates for the group with 'unknown and other' data were similar to rates for all cases and low-risk characteristics (data not shown).

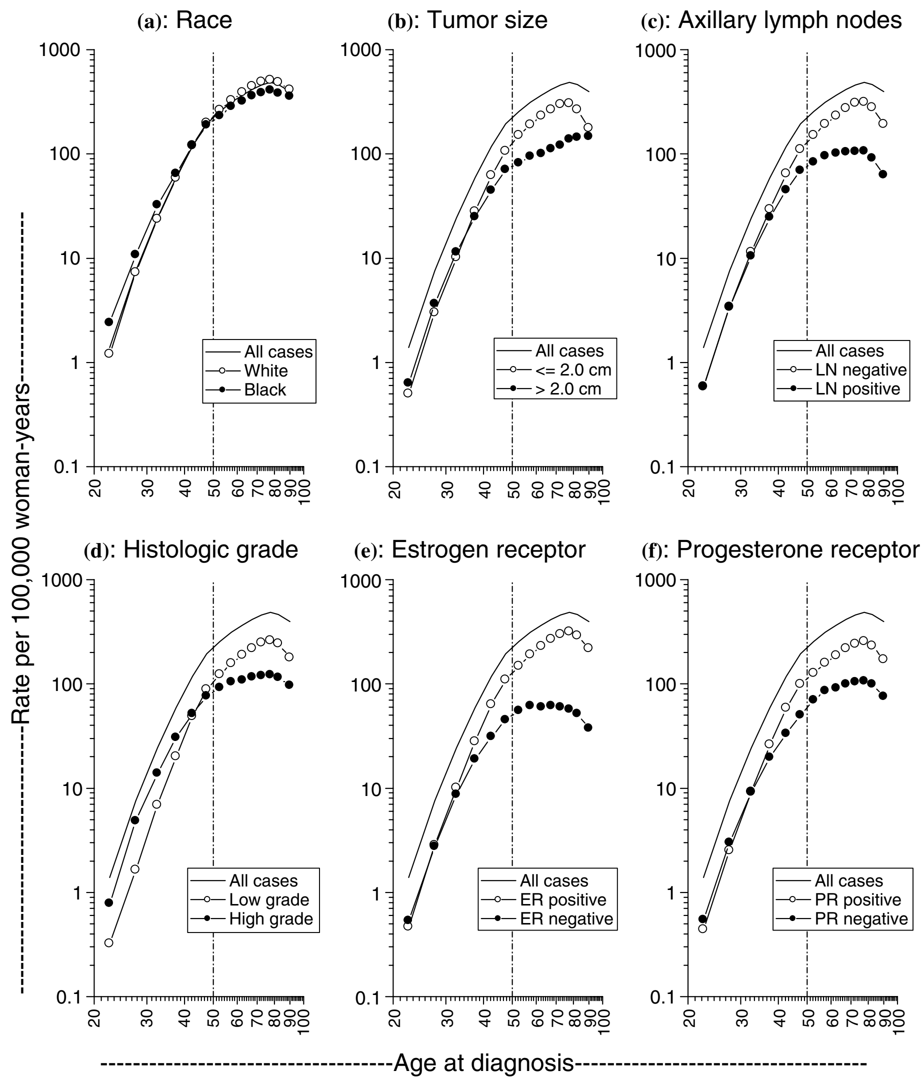

Figure 1. Age-specific incidence rates in SEER's 12 Registry Database among female breast cancer cases collected during the years 1992–2001. Each chart includes three graphs; (1) for all breast cancer cases combined, (2) for breast cancer cases defined by single low-risk tumor characteristics, and (3) for breast cancer cases defined by single high-risk tumor characteristics. (a) White (low-risk) and Black (high-risk) race; (B) Tumor size ≤ 2.0 cm (low-risk) and > 2.0 cm (high-risk); (c) Axillary lymph nodes (LN) negative (low-risk) and LN positive (high-risk); (d) Histologic grade low (low-risk) and high (high-risk); (e) Estrogen receptor (ER) positive (low-risk) and ER negative (high-risk); (f)Progesterone receptor (PR) positive (low-risk) and PR negative (high-risk).

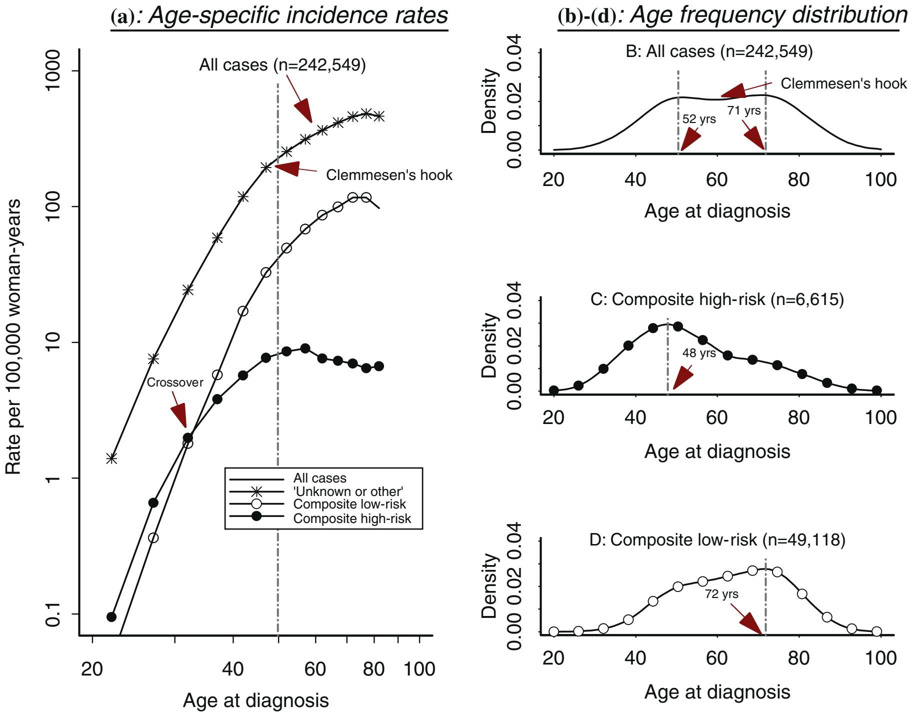

Figure 2. Age distribution at diagnosis in SEER's 12 Registry database among female breast cancer cases collected during the years 1992–2001. (a) Age-specific incidence rates for all breast cancer cases combined ($n = 242{,}549$), 'unknown and other' ($n = 186{,}816$), composite low-risk ($n = 49{,}118$), and composite high-risk ($n = 6{,}615$) prognostic profiles. Composite low-risk profiles included only those breast cancer cases among women with tumors ≤2.0 cm + axillary LN negative + low histologic grade + ER positive expression. Composite high-risk profiles included only those cases among women with tumors >2.0 cm + axillary LN positive + high histologic grade + ER negative expression. (b) Age frequency density plot among all breast cancer cases combined ($n = 242{,}549$), (c) Age frequency density plot among those breast cancer cases defined by composite high-risk profiles ($n = 6{,}615$). (D) Age frequency density plot among those breast cancer cases defined by composite low-risk profiles ($n = 49{,}118$).

Figure 2 displays age-specific incidence rates and age frequency distributions for composite tumor characteristics. Rates were higher for composite high-risk than for composite low-risk tumors until ages 30–44 years, after which there was crossover with rates higher for composite low-risk than for composite high-risk tumors (Figure 2(a)). Rates for composite high-risk tumors increased rapidly until age 50 years then plateaued and subsequently fell. Rates for composite low-risk tumors increased rapidly until age 50 years then continued to rise at a slower pace, similar to rates for all cases. Rates for breast cancers with 'unknown or other' tumor characteristics for the composite group were virtually superimposable and indistinguishable from the rate curve for all cases.

Age frequency distribution for all breast cancer cases combined had bimodal early-onset and late-onset peak frequencies at ages 52 and 71 years, respectively (Figure 2(b)). The age frequency distribution for composite high-risk tumors was unimodal with an early-onset peak frequency at age 48 years (Figure 2(c)). On the other hand, the age frequency distribution for composite low-risk tumors was predominantly unimodal with a late-onset peak frequency at age 72 years (Figure 2(d)).

Figure 3 shows age frequency distributions for all single high-risk and low-risk tumor characteristics. As for composite tumors (Figures 2(c–d)), age frequency distributions for all single high-risk tumors were predominantly unimodal with early-onset peak frequencies. On the other hand, age frequency distributions for all single low-risk tumors were predominantly unimodal with late-onset peak frequencies.

Figure 4 superimposes incidence and prognostic patterns for tumors defined by high-risk and low-risk characteristics. High-risk tumors with flattened age-specific incidence rate curves had significantly worse actuarial breast cancer survival than did low-risk tumors (Figures 4(a–b), $p < 0.001$). With 120 months of follow-up, median duration of survival was 28 months for composite high-risk compared to 41 months for composite low-risk breast cancers. Cox proportional analysis confirmed hazard ratios of 28.91 for composite high-risk compared to composite low-risk profiles (Table 2). Twenty-four months following breast cancer diagnosis, the hazard function for breast cancer death demonstrated a sharp spike for composite high-risk tumors (Figure 4(c)). This 2-year hazard peak was absent for low-risk tumors. Approximately 96 months following breast cancer diagnosis, the hazard function crossed over with low-risk tumors having slightly greater hazard rates than high-risk tumors.

Figure 5 shows hazard function for all single high-risk and low-risk tumor characteristics. All high-risk tumors had a very characteristic 2-year hazard spike, which was not present for all low-risk tumors. Hazard

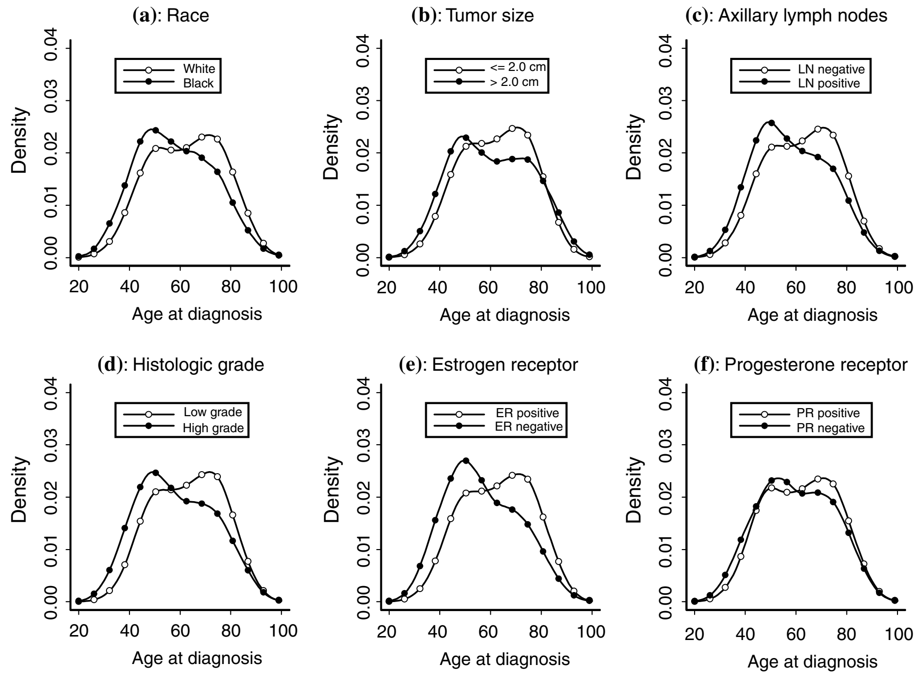

Figure 3. Age frequency density plot in SEER's 12 Registry Database among female breast cancer cases collected during the years 1992–2001. Each chart includes two graphs; (1) for breast cancer cases defined by single low-risk tumor characteristics, and (2) for breast cancer cases defined by single high-risk tumor characteristics. (a) White (low-risk) and Black (high-risk) race; Tumor size ≤2.0 cm (low-risk) and > 2.0 cm (high-risk); (c) Axillary lymph nodes (LN) negative (low-risk) and LN positive (high-risk); (d) Histologic grade low (low-risk) and high (high-risk); (e) Estrogen receptor (ER) positive (low-risk) and ER negative (high-risk); (f) Progesterone receptor (PR) positive (low-risk) and PR negative (high-risk).

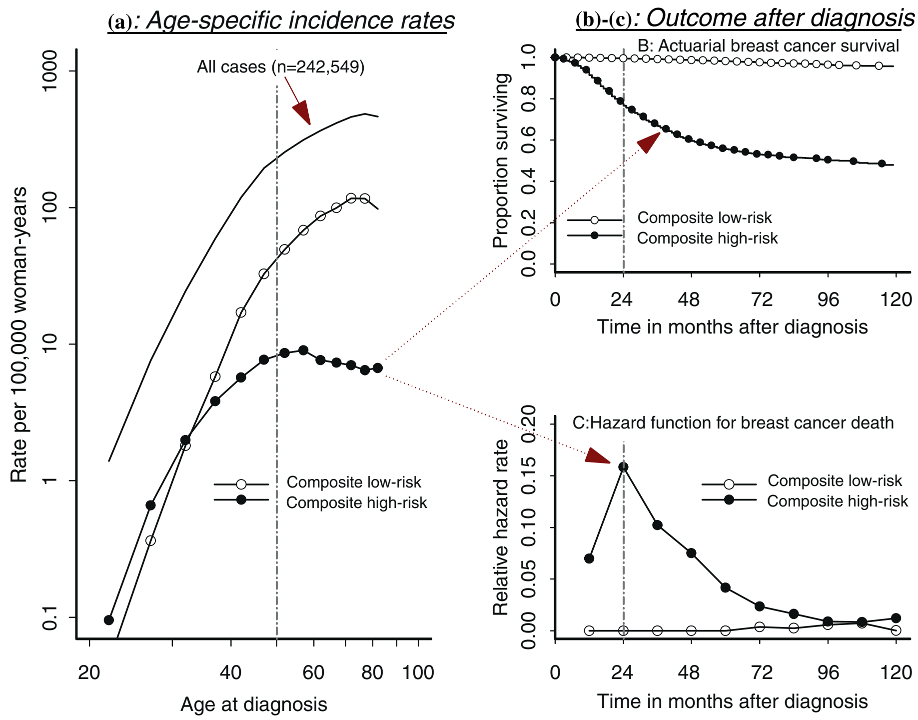

Figure 4. Breast cancer incidence and outcome in SEER's 12 Registry database among female breast cancer cases collected during the years 1992–2001. (a) Age-specific incidence rates for all breast cancer cases combined ($n = 242{,}549$), 'unknown and other' ($n = 186{,}816$), composite low-risk ($n = 49{,}118$), and composite high-risk ($n = 6{,}615$) prognostic profiles. Composite low-risk profiles included only those breast cancer cases among women with tumors ≤ 2.0 cm + axillary LN negative + low histologic grade + ER positive expression. Composite high-risk profiles included only those cases among women with tumors > 2.0 cm + axillary LN positive + high histologic grade + ER negative expression. (b) Actuarial breast cancer survival among those breast cancer cases defined by composite low-risk and composite high-risk prognostic profiles. (c) Hazard function for breast cancer death among those breast cancer cases defined by composite low-risk and composite high-risk prognostic profiles.

Table 2. Hazard ratios for standard tumor characteristics in SEER's 12 registry Database among female breast cancer cases collected during 1992–2001

Variable	All cases combined		Women < 50 years		Women ≥ 50 years	
	HR	SE	HR	SE	HR	SE
Age (years)						
< 50	1.00					
≥ 50	1.06	0.01				
Race						
White	1.00		1.00		1.00	
Black	1.99	0.02	2.22	0.03	1.88	0.02
Tumor size						
≤ 2.0 cm	1.00		1.00		1.00	
> 2.0 cm	4.37	0.01	3.32	0.03	4.80	0.02
Lymph nodes (LN)						
LN negative	1.00		1.00		1.00	
LN positive	4.79	0.02	4.01	0.03	5.11	0.02
Histologic grade						
Low	1.00		1.00		1.00	
High	3.35	0.02	3.07	0.03	3.45	0.02
ER						
ER positive	1.00		1.00		1.00	
ER negative	2.64	0.01	2.48	0.03	2.72	0.02
PR						
PR positive	1.00		1.00		1.00	
PR negative	2.37	0.01	2.40	0.03	2.34	0.02
Composite profile						
Low-risk	1.00		1.00		1.00	
High-risk	28.91	0.04	28.20	0.11	32.35	0.05

Key: HR, hazard ratio for breast cancer death, where a given characteristic is compared to a referent characteristic with an assigned HR of 1.0; SE, standard error; ER, estrogen receptor; PR, progesterone recepor; Composite low-risk prognostic profile, breast cancer cases with tumors ≤ 2.0 cm + LN negative + low grade + ER positive; Composite high-risk prognostic profile, breast cancer cases with tumors > 2.0 cm + LN positive + high grade + ER negative.

function crossed over for race, grade, ER and PR expression, but not for tumor size and LN status.

Hazard ratios (HR) for all combinations of low-risk and high-risk tumor characteristics during the time period 1992–2001 are shown in Table 2. In general, the HR for high-risk compared to low-risk tumor characteristics were greater for older than for younger women. For example, HR = 32.35 for composite high-risk profiles among women age ≥ 50 years compared to HR = 28.20 among women age < 50 years. In fact, except for PR expression, the HR for all high-risk tumor characteristics compared to low-risk tumor characteristics were greater for older than for younger women.

Discussion

Age-specific incidence rates among women with invasive breast cancer increased rapidly until age 50 years, and then diverged for high-risk and low-risk tumor characteristics (Figure 1). Yasui and Potter first demonstrated divergent age-specific incidence rate patterns for ER negative versus ER positive tumors in the Danish Breast Cancer Cooperative Group [18], which was subsequently confirmed in the SEER database [13, 19]. Similarly shaped age-specific rate curves for all single high-risk versus low-risk tumor characteristics illustrated that this peculiar phenomenon was not specific for ER expression.

Given similar incidence patterns for all single tumor characteristics, we further explored age distribution at diagnosis with composite groups containing only high-risk or low-risk tumor features (Table 1, Figure 2). Whereas age-specific incidence rates for most epithelial tumors increase steadily with aging [20, 21], rates for breast cancer overall (n = 242,549) increase more slowly after menopause (age 50 years, Figure 2(a)). The menopausal bend in age-specific rates has been termed Clemmesen's hook [22]. This unusual breast cancer incidence rate pattern with its peculiar menopausal hook

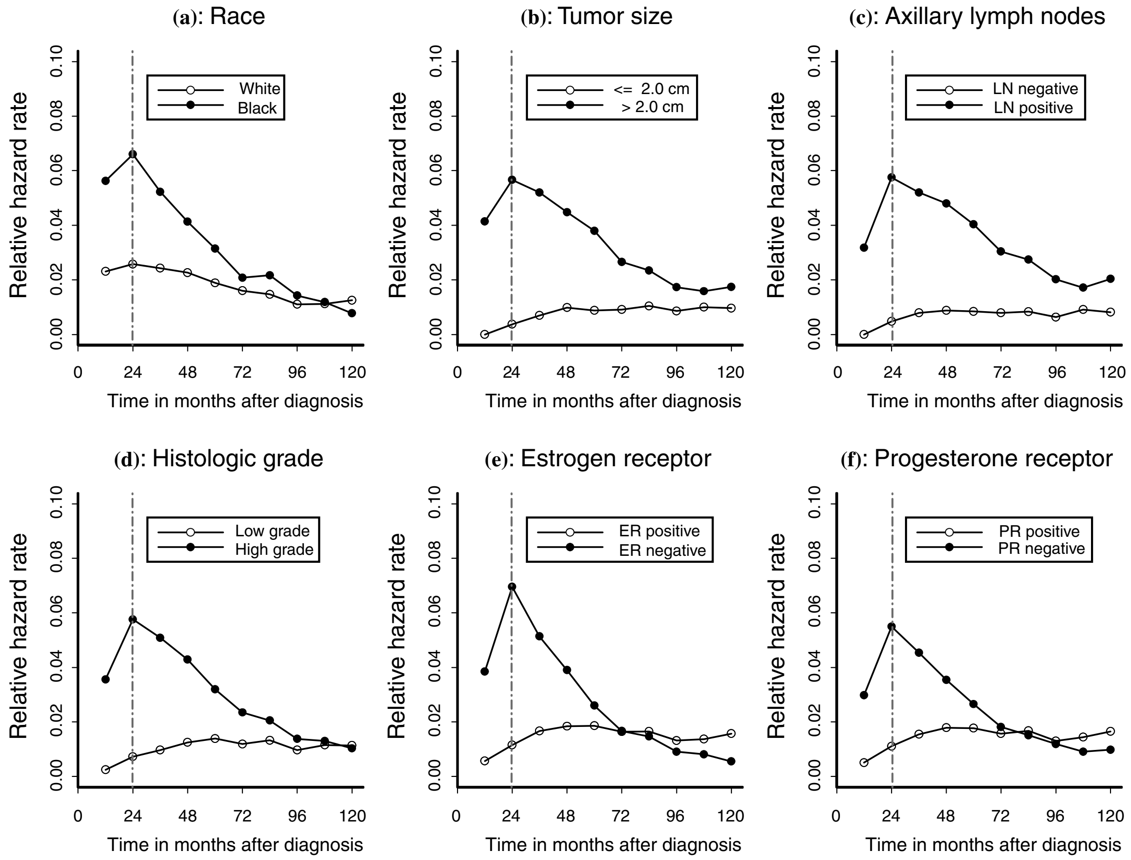

Figure 5. Hazard function in SEER's 12 Registry Database among female breast cancer cases collected during the years 1992–2001. Each chart includes two graphs; (1) for breast cancer cases defined by single low-risk tumor characteristics, and (2) for breast cancer cases defined by single high-risk tumor characteristics. (a) White (low-risk) and Black (high-risk) race; (b) Tumor size ≤ 2.0 cm (low-risk) and > 2.0 cm (high-risk); (c) Axillary lymph nodes (LN) negative (low-risk) and LN positive (high-risk); (d) Histologic grade low (low-risk) and high (high-risk); (e) Estrogen receptor (ER) positive (low-risk) and ER negative (high-risk); (f) Progesterone receptor (PR) positive (low-risk) and PR negative (high-risk).

reflects bimodal age frequency distribution (Figures 2(a–b)). More than 50 years ago [23], Elving Anderson also demonstrated that Clemmesen's hook in rates corresponded to the dip between the bimodal peaks of the age frequency distribution plot.

Figures 2(b–d) show that bimodal age frequency distribution for breast cancer overall (Figure 2(b)) reflects the summation of early-onset high-risk (Figure 2(c)) and late-onset low-risk (Figure 2(d)) unimodal breast cancer populations, mixed or juxtaposed within the general population [13, 24]. Similar early-onset and late-onset age frequency distributions existed for all single high-risk and low-risk tumor characteristics, respectively (Figure 3). Notably, age frequency distribution for male breast cancer overall is unimodal rather than bimodal [25], confirming that bimodal breast cancer overall is a female-specific or menopausal-related phenomenon [26].

Distinct incidence rate patterns predicted distinct prognostic patterns (Figure 4(a–c)). Rates for composite high-risk tumors with postmenopausally flattened age-specific rate curves had worse actuarial survival and a sharp 2-year hazard peak. The hazard ratio for composite high-risk compared to composite low-risk tumors was approximately 30 to 1 (Table 2), demonstrating an enormous relative risk for breast cancer death. In contrast, rates for composite low-risk tumors had continuously rising age-specific rate curves with improved actuarial survival and no hazard peak. Figure 5 demonstrates that all single high-risk tumors had the 2-year hazard spike, whereas all single low-risk tumors lacked this peak. Curiously, hazard rates crossed over for those characteristics related to tumor biology, i.e., race, grade, ER and PR expression, but not for those characteristics related to staging, i.e., tumor size, and LN status.

It is well-known that high-risk breast cancers tend to develop in younger rather than in older women [27, 28]. It is also known that the greatest risk for breast cancer death overall occurs approximately 2–3 years following primary breast cancer diagnosis [29]. However, to our knowledge, no one has previously linked age-specific incidence rate and hazard rate patterns, showing that the overall 2–3 year hazard peak is solely a phenomenon of early-onset high-risk tumors and completely absent for late-onset low-risk tumors.

Distinct population-based incidence and prognostic patterns in SEER appear to complement recent gene discovery, where distinct genotypes have been associated with high-risk and low-risk breast cancer phenotypes [5, 30–33]. These high-risk and low-risk incidence and gene-expression patterns are seemingly established prior to primary breast cancer diagnosis, observations that are difficult to reconcile with the commonly held view of tumor evolution from low-risk to high-risk tumor features [6, 12, 34]. Alternatively, high-risk and low-risk tumor characteristics reflect two different types of breast cancer, resulting from two risk factor profiles [13, 18] and/or two stem cells of origin [35–40].

The strength of the SEER database is its large-scale population-based design. Limitations include missing data for menopausal status, reproductive risk factors, treatment, or other factors that might impact results. However, age 50 years is a reasonable surrogate for menopause [8]. There were no reproductive risk factor data, but SEER did accurately record the most important risk factor for female breast cancer, e.g., age at diagnosis. Treatment records were unavailable, but standard tumor characteristics should predict prognosis irrespective of treatment [41]. Effective therapy might reduce the relative risk or odds of breast cancer death (i.e., the amplitude of the hazard rate), but the timing of the hazard peak for high-risk tumors always seems to occur approximately 2 years following diagnosis of primary breast cancer [29]. The presence or absence of the hazard peak also appeared to be linked to the shape of the age-specific incidence rate curve *at diagnosis*, which would not be affected by treatment *after diagnosis*. Finally, there also was a large amount of unknown data for certain tumor characteristics (Table 1). However, there was no evidence that unknown data impacted results. As in our previous SEER studies [25, 42, 43], incidence rate patterns for cases with unknown data resembled those for all breast cancer cases combined.

Notwithstanding the above caveats and despite the general view of breast cancer as a heterogeneous and chronic disease process with stepwise evolution from low-risk to high-risk tumor features [44], unique incidence and prognostic patterns in the SEER program provide population-based evidence favoring distinct high-risk and low-risk breast cancer types. Clearly, further analytic studies are required to better decipher the complicated relationship between breast cancer incidence and prognosis. This effort is warranted given the paradigmatic implications for two distinct and possibly genetically predetermined breast cancer phenotypes.

References

1. Hilsenbeck SG, Ravdin PM, de Moor CA, Chamness GC, Osborne CK, Clark GM: Time-dependence of hazard ratios for prognostic factors in primary breast cancer. Breast Cancer Res Treat 52(3): 227–237, 1998
2. McKay MJ, Langlands AO: Prognostic factors in breast cancer. New England J Med 327(18): 1317, 1992
3. Devitt JE: The significance of regional lymph node metastases in breast carcinoma. Can Med Assoc J 93: 289–293, 1965
4. MacDonald I: Biological predeterminism in human cancer. Surg Gynecol Obstet 92: 443–452, 1951
5. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al.: A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347(25): 1999–2009, 2002
6. Bernards R, Weinberg RA: Metastasis genes: A progression puzzle. Nature 418(6900): 823, 2002
7. SEER. Surveillance, Epidemiology, and End Results (SEER) Program. Public-Use Database (1973–2001), National Cancer Institute, DCCPS, Surveillence Research Program, Cancer Statistics Branch, based on the November 2003 submission. 2004; www.seer.cancer.gov
8. Morabia A, Flandre P: Misclassification bias related to definition of menopausal status in case–control studies of breast cancer. Int J Epidemiol 21(2): 222–228, 1992
9. Miettinen O, Nurminen M: Comparative analysis of two rates. Stat Med 4(2): 213–226, 1985
10. Armitage P, Doll R: The age distribution of cancer and a multistage theory of carcinogenesis. Br J Cancer 8: 1–12, 1954
11. Armitage P, Doll R: A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. Br J Cancer 11(2): 161–169, 1957
12. Anderson WF, Chu KC, Chatterjee N, Brawley OW, Brinton LA: Tumor variants by hormone receptor expression in white patients with node-negative breast cancer from the Surveillance, Epidemiology, and End Results database. J Clin Oncol 19(1): 18–27, 2001
13. Anderson WF, Chatterjee N, Ershler WB, Brawley OW: Estrogen receptor breast cancer phenotypes in the surveillance, epidemiology, and end results database. Breast Cancer Res Treat 76(1): 27–36, 2002
14. Kaplan EL, Meier P: Nonparametric estimation from incomplete observations. J Am Stat Assoc 53: 457–481, 1958
15. Peto R, Peto J: Asymptomatically efficient rank invariant test procedures. J R Stat Soc A 135(A): 185–198, 1972
16. Cox DR: Regression models and life-tables. J R Stat Soc B 34: 187–220, 1972
17. Brinton LA, Benichou J, Gammon MD, Brogan DR, Coates R, Schoenberg JB: Ethnicity and variation in breast cancer incidence. Int J Cancer 73(3): 349–355, 1997
18. Yasui Y, Potter JD: The shape of age-incidence curves of female breast cancer by hormone-receptor status. Cancer Causes Control 10(5): 431–437, 1999
19. Tarone RE, Chu KC: The greater impact of menopause on ER- than ER+ breast cancer incidence: a possible explanation (United States). Cancer Causes Control 13: 7–14, 2002
20. Lilienfeld AM, Johnson EA: The age distribution in female breast and genital cancers. Cancer 8: 875–882, 1955
21. Knudson AG: Two genetic hits (more or less) to cancer. Nat Rev Cancer 1(2): 157–62, 2001
22. Clemmesen J: Carcinoma of the breast. Br J Radiol 21(252): 583–590, 1948
23. Anderson E: Possible relationship between menopause and age at onset of breast cancer. Cancer 3: 410–411, 1950
24. Anderson WF: Puberty and genetic susceptibility to breast cancer (letter). N Engl J Med 349(11): 1088–1089, 2003
25. Anderson WF, Althuis MD, Brinton LA, Devesa SS: Is male breast cancer similar or different than female breast cancer? Breast Cancer Res Treat 83: 77–86, 2004
26. Thomas DB: Breast cancer in men. Epidemiol Rev 15(1): 220–231, 1993
27. Albain KS, Allred DC, Clark GM: Breast cancer outcome and predictors of outcome: are there age differentials? J Nat Cancer Inst Monograph 16: 35–42, 1994
28. Diab SG, Elledge RM, Clark GM: Tumor characteristics and clinical outcome of elderly women with breast cancer. J Natl Cancer Inst 92(7): 550–556, 2000
29. Jatoi I, Tsimelzon A, Weiss H, Clark GM, Hilsenbeck SG: Hazard rates of recurrence following diagnosis of primary breast cancer. Breast Cancer Research and Treatment 2004 89: 173–178, 2005.
30. Ahr A, Karn T, Solbach C, Seiter T, Strebhardt K, Holtrich U, et al.: Identification of high risk breast-cancer patients by gene expression profiling. Lancet 359(9301): 131–132, 2002
31. Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, et al.: Gene expression predictors of breast cancer outcomes. Lancet 361(9369): 1590–1596, 2003

32. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, et al.: Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 100(18): 10393–10398, 2003
33. West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, et al.: Predicting the clinical status of human breast cancer by using gene expression profiles. Proc Natl Acad Sci U S A 98(20): 11462–11467, 2001
34. Easton D: Breast cancer-not just whether but when? Nat Genet 26(4): 390–391, 2000
35. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al.: Molecular portraits of human breast tumours. Nature 406(6797): 747–752, 2000
36. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al.: Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100(14): 8418–8423, 2003
37. Boecker W, Buerger H: Evidence of progenitor cells of glandular and myoepithelial cell lineages in the human adult female breast epithelium: a new progenitor (adult stem) cell concept. Cell Prolif 36(1 Suppl): 73–84, 2003
38. Jones C, Nonni AV, Fulford L, Merrett S, Chaggar R, Eusebi V, et al.: CGH analysis of ductal carcinoma of the breast with basaloid/myoepithelial cell differentiation. Br J Cancer 85(3): 422–427, 2001
39. Foulkes WD, Brunet JS, Stefansson IM, Straume O, Chappuis PO, Begin LR, et al.: The prognostic implication of the basal-like (cyclin E high/p27 low/p53+/glomeruloid-microvascular-proliferation+) phenotype of BRCA1-related breast cancer. Cancer Res 64(3): 830–835, 2004
40. Kesse-Adu R, Shousha S: Myoepithelial markers are expressed in at least 29% of oestrogen receptor negative invasive breast carcinoma. Mod Pathol 17(6): 646–652, 2004
41. Henderson IC, Patek AJ. The relationship between prognostic and predictive factors in the management of breast cancer. Breast Cancer Res Treat 52(1–3): 261–288, 1998
42. Anderson WF, Chu KC, Chang S, Sherman ME: Comparison of age-specific incidence rate patterns for different histopathologic types of breast carcinoma. Cancer Epidemiol, Biomarkers Prev 13(7): 1128–1135, 2004
43. Anderson WF, Chu KC, Devesa SS: Distinct incidence patterns among in-situ and invasive breast carcinomas, with possible etiologic implications. Breast Cancer Res Treat 88: 149–159, 2004
44. Hellman S:Natural history of small breast cancers. J Clin Oncol 12(10): 2229–2234, 1994

Address for offprints and correspondence: William F. Anderson, DHHS/NIH/NCL/Division of Cancer Prevention, EPN Suite 2144, 6130 Executive Blvd, Bethesda, MD 20892, USA; *Tel*.: +301-594-7672; *Fax*: +301-435-6344; *E-mail*: wanderso @mail.nih.gov